

Application Design of Dengue Hemorrhagic Fever Patients Screening Using Naive Bayes Method

Cindy Astelia Ramadhan Suparman¹, Endah Purwanti¹,
Prihartini Widiyanti^{1,2*}

¹Biomedical Engineering Program, Department of Physics,
Faculty of Science and Technology, Universitas Airlangga

²Institute of Tropical Disease, Universitas Airlangga

*Corresponding author: pwidiyanti@fst.unair.ac.id

Keywords: dengue hemorrhagic fever; Naive Bayes Method; dengue patients; Medical Records; *MatlabR2013*

Abstract. Dengue Hemorrhagic Fever is a disease which is endemic in most districts / cities still becomes a public health problem in Indonesia. The awareness of people to the dengue viral infection and its symptoms are needed to decrease the fatality of this disease. The community need to be known the symptoms thereby they could intervened and prevent from falling in to worse condition. This study was conducted to design system which could diagnose the onset of the disease with 3 levels of possibilities namely Grade 1 Dengue Hemorrhagic Fever, Grade 2 Dengue Hemorrhagic Fever, and Non Dengue Hemorrhagic Fever. The system is build based on patient medical records of Dr. Wahidin Sudiro Husodo General Hospital, Mojokerto, East Java using the Naive Bayes method. The method of this study including several steps such as collecting data, preprocessing data, designing database, interface design, calculation and processing data, classification and analyzing data and evaluating application. Determining the results of the application diagnose requires posterior calculation which searches the highest values in three degrees as the results of the initial diagnose. The application as a device for an early diagnosis of dengue hemorrhagic fever has a high accuracy value of 97% out of the 30 tested data. The homogenization of the training data and the test data by sex and age can be considered in future research.

Introduction

Dengue Hemorrhagic Fever (henceforth DHF) is an infectious disease caused by infection of the dengue virus which has the fastest spread in the world through the bites of *Aedes aegypti* and *Ae. Albopictus* [1][2]. It is one of the diseases that still becomes a public health problem and is endemic in most districts / cities in Indonesia. Almost every year, the disease which is considered as an Extraordinary Event usually occurs during the rainy season in some areas. It mostly infects children aged <15 years as well as adults. The number of DHF cases in 2010 was 156.086 with 1,358 people died because of the disease [3].

The dengue virus has four serotypes, namely Den 1, Den 2, Den 3, and Den 4 which are included in group B Arbovirus group together with *Yellow Fever*, *West Nile* and *Japanese encephalitis*. There is a cross-reaction among dengue viruses in the laboratory tests and all of the four types of the viruses can cause any diseases with similar symptoms. Having virus isolation is used to get a definite diagnosis of the disease. However, it is difficult to conduct the isolation as it takes time, needs expensive special equipment, and performs the sampling on the stage of viremia [4].

The symptom of bleeding in DHF may not occur in all cases of DHF. The lightest bleeding occurs in a positive tourniquet test (Rumple Leede), which means that there is an increase in capillary fragility. Indeed the positive test may also occur in viral diseases (measles, Chikungunya fever), bacterial infections (abdominal typhus) and others [4]. The higher incidence of DHF, which is based on the reported cases, deduces that dengue cases need more serious treatment. Serious

treatment can be done by performing a diagnosis or early detection so that the incidence will be handled more quickly, which ultimately can reduce the number of cases to death. There has been much research done on the disease detection by using artificial intelligence in the modern era. This study is expected to help prospective DHF patients for an initial diagnosis to keep them vigilant in taking further actions so that the disease does not get worse [5].

There is a classification method namely *Naïve Bayes* which can be used to categorize the data on a particular class of information obtained from a set of data. The concept used in the method is a probability that predicts the future probability based on the past experience in the form of statistical models.

Based on the background described above, the researcher is interested in making an application of initial DHF diagnosis by using the *Naïve Bayes* Method. The data are obtained from the patients' medical records. This study uses and accumulates some clinical symptoms of DHF patients with the *Naïve Bayes* method which diagnoses whether or not a person suffers from the disease and determines the degrees of severity. There are clinical symptoms that are used to detect the disease such as duration of fever, muscle aches, decreased consciousness, heartburn, nausea, vomit, decreased appetite, nosebleeds, weak or impalpable pulses, and other clinical symptoms. This research is expected to help the early diagnosed DHF patients to immediately take the necessary actions, such as checking into the hospital, in order to keep their disease from getting worse.

Materials and Methods

The research was conducted in the Laboratory of Medical Instrumentation, Biomedical Engineering Program, Faculty of Science and Technology, Universitas Airlangga, as well as in Dr. Wahidin Sudiro Husodo General Hospital, East Java in about 6 months from February to June 2016.

Data for designing the early diagnosis application were medical records of patients with DHF in Dr. Wahidin Sudiro Husodo General Hospital. There were 31 parameters needed in medical records namely gender, petechiae, rash or bruising, ecchymosis, bleeding conjunctiva, nosebleeds, bleeding gums, vomit, melena, blood urine, cold tiptoes and hands, bluish skin, palpable pulses, muscle aches, decreased appetite, weakness, nausea, abdominal pain, diarrhea, constipation, loss of consciousness, age, fever, blood pressure and the difference in pulse pressure.

The procedures of the research entitled "Designing An Application of The Initial Diagnosis of Dengue Hemorrhagic Fever Patients by using *Naïve Bayes* Method" were done in several steps. Figure 1 shows a flowchart of the research methods.

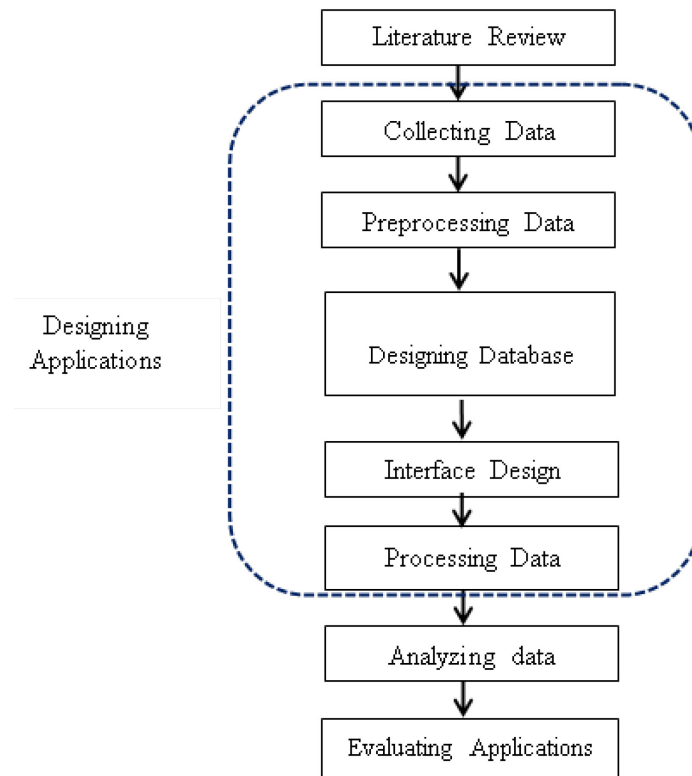


Figure 1. Flowchart of the Research Methods

The Naive Bayes method has been widely used in many cases requiring classification so that there were several steps taken to classify the processes of training and testing data by using Naïve Bayes algorithm. Figure 2 shows the flow diagram of the application of DHF initial diagnosis by using Naive Bayes.

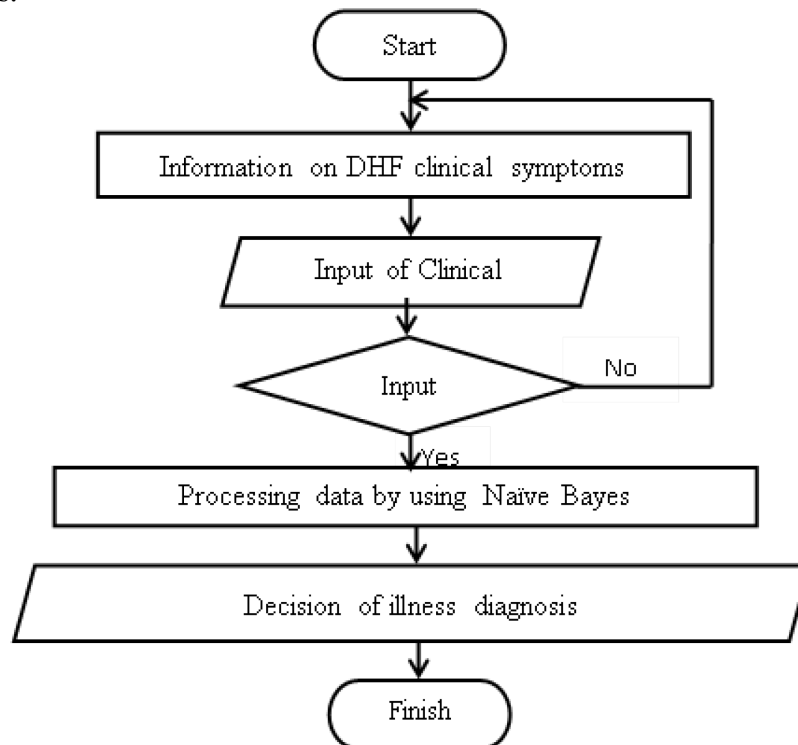


Figure 2. Flow Chart of Program

Naive Bayes Classifier

The statistical model which also includes the concepts of probability is one of the reliable models of decision support. Indeed *Naive Bayesian Classification* (NBC) is one of the methods which applies the concepts of probability. In this method, all attributes contribute with the same

degree of importance in making decision and they are independent of each other. A probability value has assumed as independent variables X_i which will merge into a special variable is the variable C . Equation combination of these two variables is shown in Equation 1 [6]

$$P(C, X_1, \dots, X_n) = P(C) \prod_{i=1}^n P(x_i | C) \quad (1)$$

The Naïve Bayes model has an independent variable which is assumed as X_i and merged into a special variable C . The equation of combination between the two variables is shown in equation 2.1 [6].

The Naïve Bayes algorithm is one of the algorithms in classification techniques. The algorithm, which is proposed a British scientist Thomas Bayes, is classification applied by using probability and statistics. It predicts future opportunities based on the earlier experience and is known as Bayes' Theorem. The theorem combined with the *Naïve* is assumed to be in independent condition [7] [8]. In Naïve Bayes classification, it is assumed whether or not there is a specific characteristic of a class has no correlation with the characteristics of other classes [3]

The equation of Bayes' theorem is as follows :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Description of the Equation 2 :

X : Data of an unknown class

H : A Hypothesis of data X is a specific class

$P(H|X)$: Probability of the hypothesis H based on the condition X (posterior probability)

$P(H)$: Probability of the hypothesis H (prior probability)

$P(X|H)$: Probability based on the condition of hypothesis H

$P(X)$: Probability X

The flow of the Naïve Bayes method is as follows: 1. Reading the training data, 2. Calculating the amount and probability. Moreover, the following steps are taken for the numerical data such a. Looking for the mean and the standard deviation of each parameter of the numeric data, b. Finding the value of probability by counting the number of the corresponding data from the same category divided by the amount of data in that category. 3. Obtaining the values of mean, standard deviation and probability in a table [4].

Results and Discussion

Interface Design

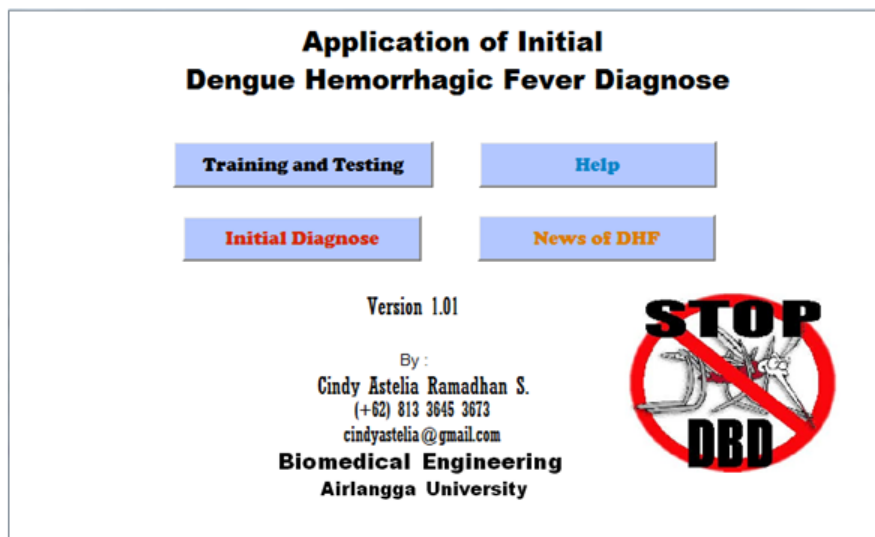


Figure 3. Main Menu Page

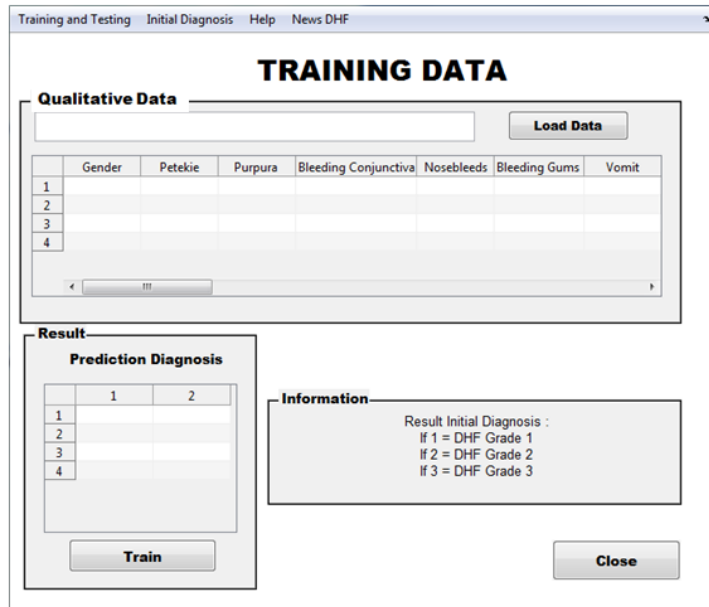


Figure 4. Training Page

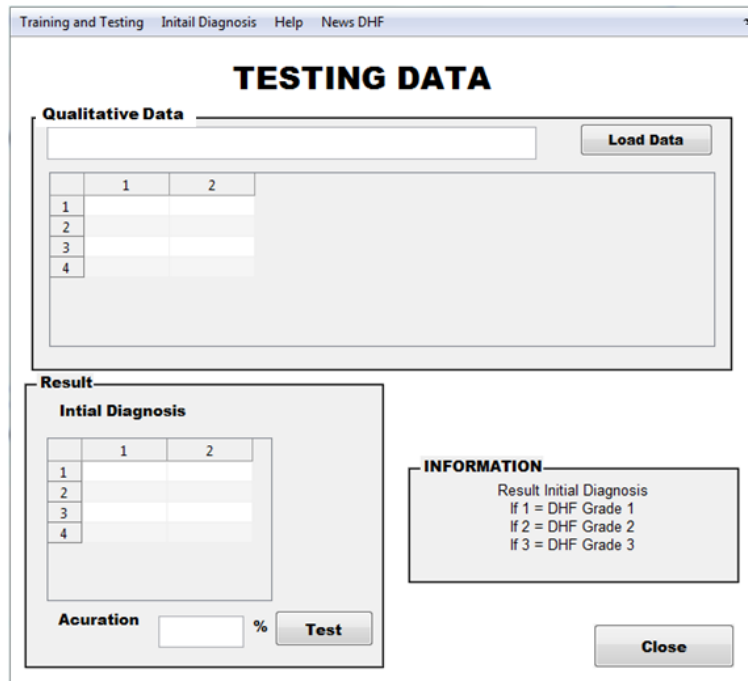


Figure 5. Testing Page

Figure 6. Initial Diagnosis Page

Calculation of Naive Bayes

a. Calculation of *Likelihood*

Calculation of *Likelihood* is used to calculate the probability occurrence of each parameter (X_i) that affects each class (C_i).

The calculation of *Likelihood* is shown in the Equation 3:

$$P(X|C) = \frac{X_i}{C_i} \quad (3)$$

The Sex Probability of DHF Grade 1 (C_1) is shown in the Equation 4:

$$P(X(\text{Sex})|C_1) = \frac{X_i}{C_1} \quad (4)$$

$$P(X(\text{Sex})|C_1) = \frac{X(\text{Male})}{C_1} = \frac{42}{65} = 0.64615$$

$$P(X(\text{Sex})|C_1) = \frac{X(\text{Female})}{C_1} = \frac{23}{65} = 0.35385$$

Of the 162 training data, it is known that there are 65 data of DHF grade. Grade 1 (C_1), 49 data of DHF Grade 2 (C_2), and 48 data of Non DBD (C_3). The prior Probability calculation of the possibility of DHF Grade 1 $P(C_1)$ by the Equation 5 is as follows:

$$P(C_i) = \frac{C_i}{C} \quad (5)$$

The prior Probability calculation of the possibility of DHF Grade 1 $P(C_1)$ by the Equation 6 is as follows:

$$P(C_1) = \frac{C_1}{C} = \frac{65}{162} = 0,4012 \quad (6)$$

The prior Probability calculation of the possibility of DHF Grade 2 $P(C_2)$ by Equation 7 is as follows:

$$P(C_2) = \frac{C_2}{C} = \frac{49}{162} = 0,3024 \quad (7)$$

The prior Probability calculation of the possibility of non DHF Grade 2 $P(C2)$ by the Equation 8 is as follows:

$$P(C2) = \frac{C2}{C} = \frac{48}{162} = 0,2926 \quad (8)$$

Classification Results and Analysis

The posterior probability calculation is required to determine the results of the classification. The calculation is done to determine chances of a hypothetical right which is the result of data classification. The posterior probability calculation on the training data is done by performing the Equations 8 and 9. It is necessary to calculate the *likelihood* of the data which is used to calculate the probability occurrence of each feature affecting the degrees prior to determine the posterior probability calculation [9].

$$\text{Probabilitas posterior} = \frac{\text{Likelihood} \cdot \text{Class Prior}}{\text{Predictor Prior}} \quad (9)$$

Or can also be written as

Probabilitas posterior : posterior probability

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (10)$$

The test is done on the 30 medical records. There are 10 patients suffering from DHF Grade 1, 10 patients enduring of DHF Grade 2, and 10 patients classifying as non-dengue grade. As stated in one of the first test data, the posterior probability calculated by the application is as follows:

$$P(C1|X) = 0.3916$$

$$P(C2|X) = 0.0243$$

$$P(C3|X) = 0.0654$$

Therefore, the decision on what application to take is based on the highest probability namely Grade 1 (C1).

Table 1. Experts' Diagnosis Results and Initial Diagnosis on the Application

No of TestData	Experts' Results	Posteriorprobability			Program Results	Information
		Grade 1	Grade 2	NonDHF		
1	Grade 1	0.3916	0.0243	0.0654	Grade 1	matched
2	Grade 1	0.3917	0.8748	0.0273	Grade 2	unmatched
3	Grade 1	0.1469	0.0937	0.9813	Grade 1	matched
4	Grade 1	0.0003	0	0	Grade 1	matched
5	Grade 1	0.0109	0.0012	0.0008	Grade 1	matched
6	Grade 1	0.0245	0.0234	0	Grade 1	matched
7	Grade 1	0.0109	0	0	Grade 1	matched
8	Grade 1	0.1679	0	0	Grade 1	matched
9	Grade 1	0.0163	0.0061	0	Grade 1	matched
10	Grade 1	0.0007	0	0	Grade 1	matched
11	Grade 2	0	0.0476	0	Grade 2	matched
12	Grade 2	0	0	0	Grade 2	matched
13	Grade 2	0	0.0003	0	Grade 2	matched
14	Grade 2	0	0.5103	0	Grade 2	matched
15	Grade 2	0	0.0875	0	Grade 2	matched
16	Grade 2	0	0.0005	0	Grade 2	matched
17	Grade 2	0	0.8748	0.0001	Grade 2	matched
18	Grade 2	0	0.0794	0	Grade 2	matched

19	Grade 2	0	0	0	Grade 2	matched
20	Grade 2	0	0.5292	0.0001	Grade 2	matched
21	Non DHF	0	0	0.0654	Non DHF	matched
22	Non DHF	0	0	0.1163	Non DHF	matched
23	Non DHF	0	0	0.0129	Non DHF	matched
24	Non DHF	0	0	0.0087	Non DHF	matched
25	Non DHF	0	0	0.0013	Non DHF	matched
26	Non DHF	0	0	0.1163	Non DHF	matched
27	Non DHF	0	0	0.0129	Non DHF	matched
28	Non DHF	0	0	0.0010	Non DHF	matched
29	Non DHF	0	0.0104	0.1472	Non DHF	matched
30	Non DHF	0	0.0005	0.3925	Non DHF	matched

Evaluation of the Application

Based on the results of the program diagnosis and the experts' diagnosis, the calculation of the application accuracy levels is as follows:

$$Accuracy = \frac{\Sigma \text{ correct data}}{n} * 100 \% \quad (11)$$

$$Accuracy = \frac{29}{30} * 100 \% = 97\%$$

Description of the Equation 9:

Σ correct data = number of the corresponding test data

N = Total Number of test data

The results of the calculation on the application evaluation indicates that the level of accuracy of the initial diagnosis device which is compared with the initial experts' diagnosis is 97%. This application could increase the awareness of people to Dengue Haemorrhagic Fever (DHF). By using this application, the condition of people suspected to DHF would not deteriorated. This program extremely advocating the prevention effort of DHF. In a disease-endemic country, it is important to organize laboratory services in the context of patients' needs and disease control strategies [10].

Conclusion

1. The Naive Bayes method could be used to classify the initial diagnosis of the DHF disease by using patients' medical records as the data.
2. The application of this initial diagnosis has quite high accuracy, namely 97%. The results of the accuracy are obtained by comparing the number of correct data which are collected from test results divided by the total number of the test data.

Acknowledgement

The author would like to thank to Rumah Sakit Umum Dr. Wahidin Sudiro Husodo for permission and technical assistance. We deliver our gratitude to all patients for their participation in this study.

References

- [1] Center for Disease Control and Prevention. US department of Health and Human Service. <https://www.cdc.gov/dengue/epidemiology/> , p 1-2
- [2] Doggett, Stephen L. 2003. *NSW Arbovirus Surveillance & Vector Monitoring Program: Mosquito Photos*. <http://medent.usyd.edu.au/arbovirus/mosquit/photos/mosquitphotos.htm>
- [3] DepKes RI, *Profil Kesehatan Indonesia 2010*, Jakarta, Kementerian Kesehatan Republik Indonesia (2011), p 61-63
- [4] Siregar, Ahmad Dian, Gambaran Pasien Demam Berdarah Dengue di Bangsal Anak RSUD Dr. Abdul Aziz Singkawang Tahun 2005, DEXA MEDIA No. 2, Vol. 19(2006), p 66-68
- [5] Ginanjar, Genis, *Demam Berdarah A Survival Guide*. Yogyakarta, Penerbit B-First (2008), p 31-34
- [6] Lowd, Daniel, Domingos, Pedro, *Naïve Bayes Moels for Probability Estimation*, Washington. Department of Computer Science and Engineering, University od Washington, 2005, p 1-3
- [7] Bustami, Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika* Vol 8, No 1 (2014), p 129-134
- [8] Kusumadewi S., *Klasifikasi Statuz Gizi Menggunakan Naïve Bayesian Classification*, *Comm IT*, Vol. 3, No. 1 (2009) 6 – 11
- [9] Suwardi U., Syukur A., Anggi P. R, *Komparasi Algoritma Backpropagation, Nearest Neighbor Dan Desicion Tree Untuk Mendeteksi Penyakit Demam Berdarah Pada Pasien Opname*, Pascasarjana Teknik Informatika, Universitas Dian Nuswantara (2012), p 59-61.
- [10] WHO, *Dengue Guidelines for Diagnosis, Treatment, Prevention and Control*, New Edition (2009), p 104